



Survival-Inferred Fragility Index of Phase 3 Clinical Trials Evaluating Immune Checkpoint Inhibitors

Bomze, David ; Asher, Nethanel ; Hasan Ali, Omar ; Flatz, Lukas ; Azoulay, Daniel ; Markel, Gal ;
Meirson, Tomer

Abstract: Importance In science and medical research, extreme and dichotomous conclusions may be drawn based on whether the P value falls above or below the threshold. The fragility index (ie, the minimum number of changes from nonevents to events resulting in loss of statistical significance) captures the vulnerability of statistics in trials with binary outcomes. There are a growing number of clinical trials of immune checkpoint inhibitors (ICIs), as well as expanding eligibility for patients to receive them. The robustness of survival outcomes in randomized clinical trials (RCTs) should be evaluated using the fragility index extended to time-to-event data. Objective To calculate the fragility of survival data in RCTs evaluating ICIs. Design, Setting, and Participants In this cross-sectional study, data on phase 3 prospective RCTs investigating ICIs included in PubMed from inception until January 1, 2020, were extracted. Two- or three-group studies reporting results for overall survival were eligible for the survival-inferred fragility index (SIFI) calculation, which is the minimum number of reassignments of the best survivors from the interventional group to the control group resulting in loss of significance (defined as $P < .05$ by log-rank test). For nonsignificant results, a negative SIFI was calculated by reversing the direction of reassignment (from the control group to the interventional group). Main Outcomes and Measures Survival-inferred fragility index. Results A total of 45 phase 3 prospective RCTs (4 of which had 3 groups, for a total of 49 groups) were identified, of which 6 (13%) investigated anti-cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) agents, 25 (56%) investigated anti-programmed cell death 1 (PD-1) agents, 12 (27%) investigated anti-programmed cell death 1 ligand 1 agents, and 3 (7%) investigated the combination of anti-CTLA-4 and anti-PD-1 agents. The median SIFI was 5 (interquartile range, -4 to 12) for the intention-to-treat analysis; for these trials, the SIFI was 1% or less of the total sample size in 17 of 49 populations (35%). In 25 of the 49 intention-to-treat populations (51%), the SIFI was less than the number of censored patients in the intervention group shortly after randomization (defined as $<5\%$ of the follow-up time). Conclusions and Relevance This study suggests that many phase 3 RCTs evaluating ICI therapies have a low SIFI for overall survival, resulting in uncertainty regarding their potential clinical benefit. Although not a definitive solution for the problems arising from dichotomization, SIFI provides an additional means of assessing and communicating the strength of statistical conclusions.

DOI: <https://doi.org/10.1001/jamanetworkopen.2020.17675>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-191200>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Bomze, David; Asher, Nethanel; Hasan Ali, Omar; Flatz, Lukas; Azoulay, Daniel; Markel, Gal; Meirson, Tomer (2020). Survival-Inferred Fragility Index of Phase 3 Clinical Trials Evaluating Immune Checkpoint Inhibitors. JAMA Network Open, 3(10):e2017675.

DOI: <https://doi.org/10.1001/jamanetworkopen.2020.17675>



Original Investigation | Oncology

Survival-Inferred Fragility Index of Phase 3 Clinical Trials Evaluating Immune Checkpoint Inhibitors

David Bomze, MSc; Nethanel Asher, MD; Omar Hasan Ali, MD; Lukas Flatz, MD; Daniel Azoulay, MD, PhD; Gal Markel, MD, PhD; Tomer Meirson, BSc

Abstract

IMPORTANCE In science and medical research, extreme and dichotomous conclusions may be drawn based on whether the *P* value falls above or below the threshold. The fragility index (ie, the minimum number of changes from nonevents to events resulting in loss of statistical significance) captures the vulnerability of statistics in trials with binary outcomes. There are a growing number of clinical trials of immune checkpoint inhibitors (ICIs), as well as expanding eligibility for patients to receive them. The robustness of survival outcomes in randomized clinical trials (RCTs) should be evaluated using the fragility index extended to time-to-event data.

OBJECTIVE To calculate the fragility of survival data in RCTs evaluating ICIs.

DESIGN, SETTING, AND PARTICIPANTS In this cross-sectional study, data on phase 3 prospective RCTs investigating ICIs included in PubMed from inception until January 1, 2020, were extracted. Two- or three-group studies reporting results for overall survival were eligible for the survival-inferred fragility index (SIFI) calculation, which is the minimum number of reassignments of the best survivors from the interventional group to the control group resulting in loss of significance (defined as $P < .05$ by log-rank test). For nonsignificant results, a negative SIFI was calculated by reversing the direction of reassignment (from the control group to the interventional group).

MAIN OUTCOMES AND MEASURES Survival-inferred fragility index.

RESULTS A total of 45 phase 3 prospective RCTs (4 of which had 3 groups, for a total of 49 groups) were identified, of which 6 (13%) investigated anti-cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) agents, 25 (56%) investigated anti-programmed cell death 1 (PD-1) agents, 12 (27%) investigated anti-programmed cell death 1 ligand 1 agents, and 3 (7%) investigated the combination of anti-CTLA-4 and anti-PD-1 agents. The median SIFI was 5 (interquartile range, –4 to 12) for the intention-to-treat analysis; for these trials, the SIFI was 1% or less of the total sample size in 17 of 49 populations (35%). In 25 of the 49 intention-to-treat populations (51%), the SIFI was less than the number of censored patients in the intervention group shortly after randomization (defined as <5% of the follow-up time).

CONCLUSIONS AND RELEVANCE This study suggests that many phase 3 RCTs evaluating ICI therapies have a low SIFI for overall survival, resulting in uncertainty regarding their potential clinical benefit. Although not a definitive solution for the problems arising from dichotomization, SIFI provides an additional means of assessing and communicating the strength of statistical conclusions.

Key Points

Question How stable are the conclusions of phase 3 randomized clinical trials of immune checkpoint inhibitors in oncology?

Findings This cross-sectional study of 45 randomized clinical trials calculated the survival-inferred fragility index and found that many oncologic trials assessing immune checkpoint inhibitors have a low survival-inferred fragility index, often less than a small fraction of the sample size and less than the number of patients censored soon after randomization.

Meaning These results challenge the robustness of many phase 3 randomized clinical trials of immune checkpoint inhibitors in oncology and address the uncertainty regarding their potential clinical benefit.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

JAMA Network Open. 2020;3(10):e2017675. doi:10.1001/jamanetworkopen.2020.17675

Open Access. This is an open access article distributed under the terms of the CC-BY License.

Introduction

Immune checkpoint inhibitors (ICIs) targeting cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) or programmed cell death 1 (PD-1) and programmed cell death 1 ligand 1 (PD-L1) have revolutionized cancer treatment and led to their approval as first-line therapies, either alone or in combination with chemotherapy, for many solid tumors and hematologic malignant neoplasms.¹ However, the clinical benefit associated with ICIs cannot be generalized into a single category, as the therapeutic effectiveness varies widely across different cancer indications.²⁻⁷ The number of active clinical trials of ICIs is growing rapidly, along with an increased pace of accelerated approvals by the US Food and Drug Administration (FDA).^{8,9} The eligibility criteria for ICI therapy are dynamic, and results of postmarketing studies often lead to label revisions, with more changes expected to follow.¹⁰ Despite the popularity of ICIs and the expanding eligibility for expensive and potentially toxic treatments, the percentage of eligible patients who benefit from ICIs is decreasing.^{10,11} This gap between ICI eligibility and clinical benefit is concerning and is not fully understood.

Since the introduction of the P value almost a century ago, reliance on a fixed cutoff serving as the gatekeeper for establishing significance in clinical trials has caused controversy.^{12,13} Statistically significant differences in outcomes using an arbitrary threshold ($P < .05$) may not be clinically relevant, especially when the estimated outcome does not offer substantial clinical benefit.^{14,15} The fragility of statistical inference can be signified by the ease with which a significant P value ($P < .05$) crosses over the significance threshold ($P > .05$).^{16,17} Johnson et al¹⁸ introduced a method to compute the fragility for survival analysis by iteratively adding artificial patients to the experimental group with events at the mean exposure time of all individuals until significance is lost. Using this method, one study has recently shown that the fragility index of time-to-event data can be used to estimate the level of confidence of positive results reported in randomized clinical trials (RCTs) leading to FDA approval of anticancer drugs.¹⁹ However, this approach that simulates average “virtual” patients might inflate the fragility estimate as patients at the extreme, who contribute the most to the survival curves, are disregarded. Many possible ways could be formulated to estimate the fragility of survival data. Therefore, we aimed to define a simple and intuitive fragility measure for survival analysis, based on real-life conditions, that captures the vulnerability of the data. Hence, we define the survival-inferred fragility index (SIFI) as the minimum number of reassignments of the best survivors (defined as the patients with the longest follow-up time, regardless of having an event or being censored; the worst survivors were defined as the patients with the earliest events) from the experimental group to the control group resulting in loss of significance (**Figure 1**). The purpose of this study is to evaluate the fragility of phase 3 RCTs comparing ICIs with control or standard treatments in a time-aware context.

Methods

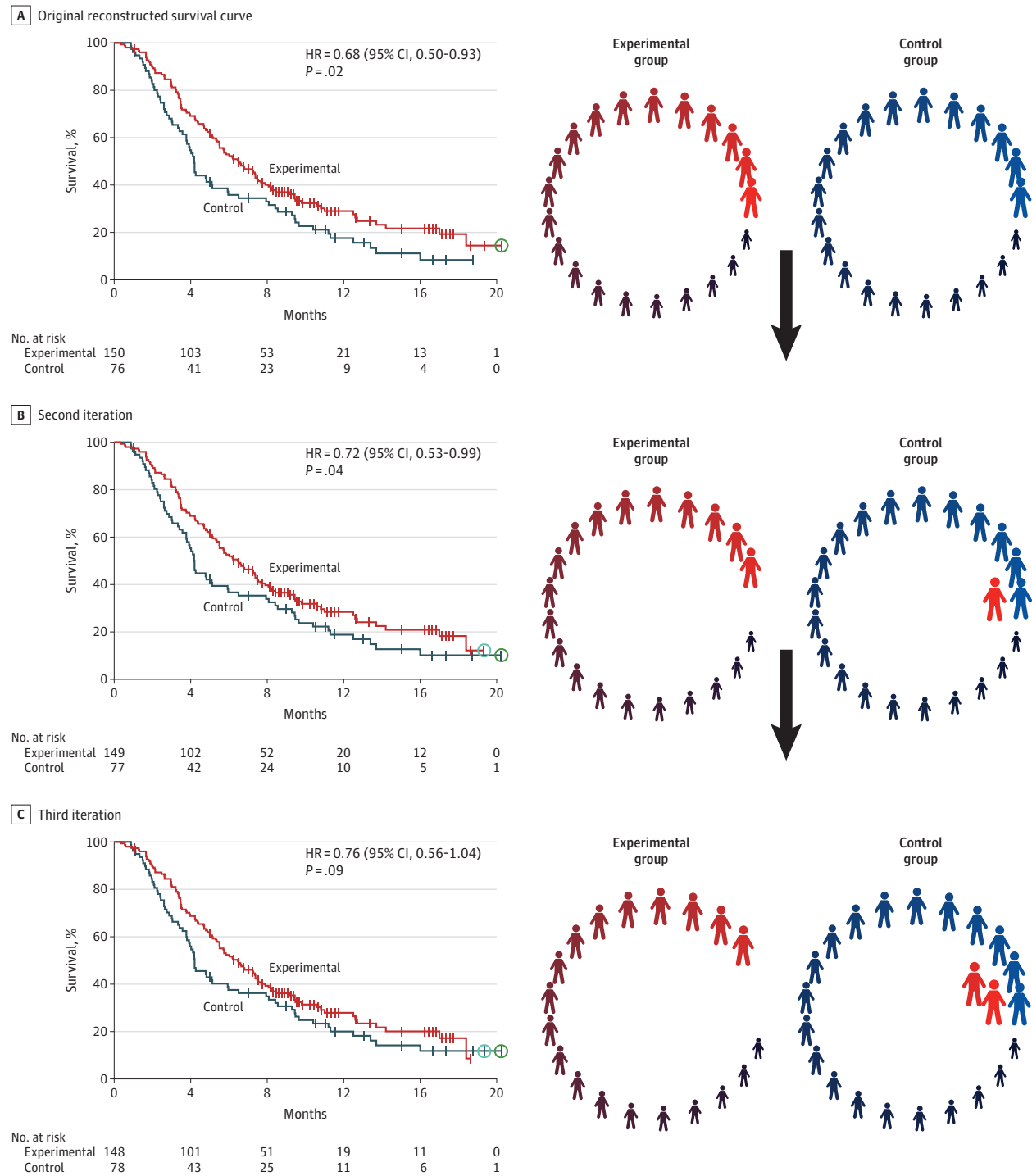
Study Design

The cross-sectional study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.²⁰ We searched PubMed from inception until January 1, 2020, for phase 3 RCTs of ICIs (anti-CTLA-4, anti-PD-1, and anti-PD-L1) compared with standard treatment in solid and hematologic malignant neoplasms. Key words for the literature search included *randomised*, *randomized*, *phase 3*, *phase III*, *ipilimumab*, *nivolumab*, *pembrolizumab*, *cemiplimab*, *durvalumab*, *avelumab*, and *atezolizumab*. For the fragility analysis, we included 2- or 3-group studies that reported overall survival as a primary or secondary outcome. We excluded retrospective studies, pooled studies, and post hoc subgroup analyses. When duplicate publications for the same trial were identified, we included the most updated publication. We abstracted information on trial design and the number of enrolled patients in the study. According to institutional review board policy, ethical approval is not required because no human data were included and publicly available information was used.

Data Extraction

Overall survival data from 45 trials were extracted from Kaplan-Meier curves in the main text using Digitizelt software (Digitizelt) and the method by Wei and Royston²¹ using Stata, version 13.0

Figure 1. Example of Survival-Inferred Fragility Index (SIFI) Calculation of Overall Survival



A, Original reconstructed survival curve. B, Second iteration of the survival curve. C, Third iteration of the survival curve. The SIFI in this example is 2, which is the iterative reassignment of the best survivors (designated by circles at the end of the survival

curves) from the experimental group to the control group, until positive significance is lost (defined as $\alpha = .05$ using log-rank test). HR indicates hazard ratio.

(StataCorp). This reverse-engineering strategy enabled us to reproduce survival time and censoring status at the individual patient level with minor differences between reconstructed and published data.¹⁹ We excluded publications of trials with raster images in which data extraction could not be performed directly. We separated the populations into 2 cohorts—the intention-to-treat (ITT) populations, which also included modified ITT populations, and subgroup populations.

Statistical Analysis

The SIFI was calculated from Kaplan-Meier curves by the iterative redesignation of the best survivors from the experimental group to the control group until positive significance (defined as $P < .05$ obtained with a 2-sided log-rank test) was lost. Negative SIFI was calculated similarly, but the direction was opposite—redesignation of the best survivors from the control group to the experimental group. In addition to the default SIFI application (flipping the best survivor from the intervention group to the control group), we defined 3 alternative approaches: flipping the worst survivor from the experimental group to the control group, cloning the best survivor in the experimental group into the control group, and cloning the worst survivor in the control group to the experimental group. P values were calculated with the 2-sided unstratified log-rank test. The follow-up time distribution was calculated using the *prodlim* package in R (R Foundation for Statistical Computing). All other analyses were performed in R, version 3.5.0. The code used to calculate SIFI is available online.²²

To provide a reference for the ranges of SIFI for various parameters of survival data, we generated synthetic survival data with the *survsim* package in R.²³ The “*simple.surv.sim*” function was used with the Weibull distribution for both the time to event and the time to censoring. The cohort size was set to range from 100 to 1200 individuals in intervals of 100 (with a 1:1 allocation). The ancillary parameter for the events was set to 1.5, and the ancillary parameter for the censoring was set to 2, 4, 6, 8, or 10. The covariate for the effect size was set to all values between -1 and 0.2 in increments of 0.05 . The β_0 parameter for the event distribution was set to 2.0 , and the β_0 for the censoring distribution was set to 2.01 .

Results

For the period until January 1, 2020, we identified 45 phase 3 RCTs (4 of which had 3 groups, for a total of 49 groups)^{2,7,24-62} evaluating ICI therapies that met the inclusion criteria for survival fragility analysis. All except 2 multiple myeloma trials (4%)^{2,47} investigated solid tumors. Six trials (13%) investigated an anti-CTLA-4 agent (ipilimumab),^{6,24-28} 25 trials (56%) investigated anti-PD-1 agents (nivolumab and pembrolizumab),^{2,3,29-51} 12 trials (27%) investigated anti-PD-L1 agents (atezolizumab, avelumab, and durvalumab),^{5,7,52-61} and 3 trials (7%) investigated the combination of anti-CTLA-4 and anti-PD-1 agents (ipilimumab and nivolumab).^{4,36,62} We could not calculate the SIFI for 2 trials (CA184-002 and CA184-043)^{63,64} because of an incompatible graphical format of the Kaplan-Meier plots. The median sample size for the eligible trials was 559 (interquartile range [IQR], 418-727). The SIFI was calculated for an additional 36 subgroups (eg, PD-L1, $\geq 1\%$) in 15 trials with a median sample size of 362 (IQR, 217-486).^{4,7,28,31,36,37,41,46,51-53,56,57,59,62}

Thirty-four of the 49 reconstructed overall survival curves in the ITT population (69%), which includes the modified ITT population, and 26 of the 36 subgroup populations (72%) were significant ($P < .05$) (Table 1).^{2,7,24-62} The median SIFI for ITT populations was 5 (IQR, -4 to 12) (ie, a median of 5 patients [among best survivors] reassigned to the control group was required to shift the results from significant to nonsignificant). The median SIFI for subgroup populations was 3.5 (IQR, 1-6.3) (eTable in the Supplement). In comparison, the fragility estimate for survival data by Johnson et al¹⁸ is unable to estimate fragility for nonsignificant results (negative fragility) and depicts higher values, with a median of 29 (IQR, 0-51) for the ITT populations and 29 (IQR, 0-43) for the subgroup populations. The absolute SIFI was less than 1% of the sample size in 17 (35%) of the 49 ITT populations and 10 (28%) of the 36 subgroup populations. Furthermore, in 25 (51%) of the 49 ITT

Table 1. SIFI of Overall Survival Calculated for 45 Phase 3 Trials Evaluating Immune Checkpoint Inhibitors in the Intention-to-Treat Populations

Intervention	Control	Tumor type	Clinical trial	Year	Sample size	HR	P value ^a	SIFI ^b
Anti-CTLA-4								
Ipilimumab + dacarbazine	Dacarbazine	Melanoma	CA184-024 ²⁴	2015	502	0.72	.001	7
Ipilimumab	Placebo	Melanoma	CA184-029 ²⁵	2016	951	0.72	.002	20
Ipilimumab	Placebo	PC	CA184-095 ⁶	2017	602	1.11	.47	-21
Ipilimumab + etoposide + platinum	Etoposide + platinum	SCLC	CA184-156 ^{26c}	2016	954	0.94	.44	-7
Ipilimumab + paclitaxel + carboplatin	Paclitaxel + carboplatin	Squamous NSCLC	CA184-104 ^{27c}	2017	749	0.91	.17	-3
Ipilimumab, 10 mg/kg	Ipilimumab, 3 mg/kg	Melanoma	CA184-169 ²⁸	2017	727	0.84	.04	1
Anti-PD-1								
Nivolumab	Docetaxel	Squamous NSCLC	CheckMate 017 ²⁹	2015	272	0.59	.0003	8
Nivolumab	Docetaxel	Nonsquamous NSCLC	CheckMate 057 ³⁰	2015	582	0.73	.004	5
Nivolumab	Everolimus	RCC	CheckMate 025 ³¹	2019	821	0.74	.003	10
Nivolumab	Platinum	NSCLC	CheckMate 026 ^{32d}	2017	423	1.02	.77	-16
Nivolumab	Placebo	GC or GEJC	ATTRACTION-2 ³³	2017	493	0.63	<.0001	10
Nivolumab	Dacarbazine or carboplatin + paclitaxel	Melanoma	CheckMate 037 ³⁴	2018	405	0.95	.59	-9
Nivolumab	Dacarbazine	Melanoma	CheckMate 066 ³⁵	2019	418	0.46	<.0001	30
Nivolumab	Ipilimumab	Melanoma	CheckMate 067 ³⁶	2018	631	0.65	<.0001	23
Nivolumab	Methotrexate, docetaxel, or cetuximab	HNSCC	CheckMate 141 ³⁷	2018	361	0.68	.001	5
Nivolumab	Paclitaxel or docetaxel	ESCC	ATTRACTION-3 ³⁸	2019	419	0.77	.015	2
Nivolumab	Docetaxel	NSCLC	CheckMate 078 ³⁹	2019	504	0.68	.003	5
Pembrolizumab	Platinum	NSCLC	KEYNOTE-024 ^{40e}	2016	305	0.60	.01	6
Every 2 wk								
Ipilimumab	Ipilimumab	Melanoma	KEYNOTE-006 ³	2017	557	0.68	.001	15
Every 3 wk								
Ipilimumab	Ipilimumab	Melanoma	KEYNOTE-006 ³	2017	555	0.68	.001	14
Pembrolizumab	Methotrexate, docetaxel, or cetuximab	HNSCC	KEYNOTE-040 ⁴¹	2018	495	0.8	.02	3
Pembrolizumab	Paclitaxel	GC or GEJC	KEYNOTE-061 ⁴²	2018	395	0.82	.06	-1
Pembrolizumab + pemtrexed + platinum	Pemtrexed + platinum	Nonsquamous NSCLC	KEYNOTE-189 ⁴³	2018	616	0.49	<.0001	40
Pembrolizumab + carboplatin + paclitaxel or nab-paclitaxel	Carboplatin + paclitaxel or nab-paclitaxel	Squamous NSCLC	KEYNOTE-407 ⁴⁴	2018	559	0.64	.002	10
Pembrolizumab	Paclitaxel, docetaxel, or vinflunine	UC	KEYNOTE-045 ⁴⁵	2019	542	0.70	.0005	9
Pembrolizumab	Cetuximab + platinum + fluorouracil	HNSCC	KEYNOTE-048 ⁴⁶	2019	601	0.83	.02	2
Pembrolizumab + platinum + fluorouracil	Cetuximab + platinum + fluorouracil	HNSCC	KEYNOTE-048 ⁴⁶	2019	559	0.77	.005	5
Pembrolizumab + pomalidomide + dexamethasone	Pomalidomide + dexamethasone	MM	KEYNOTE-183 ⁴⁷	2019	249	1.61	.14	-47
Pembrolizumab + lenalidomide + dexamethasone	Lenalidomide + dexamethasone	MM	KEYNOTE-185 ²	2019	301	2.06	.06	-79
Pembrolizumab	Placebo	HCC	KEYNOTE-240 ⁴⁸	2020	413	0.78	.04	1
Pembrolizumab + axitinib	Sunitinib	RCC	KEYNOTE-426 ⁴⁹	2019	861	0.53	.0003	40
Pembrolizumab epacadostat	Pembrolizumab	Melanoma	KEYNOTE-252 ⁵⁰	2019	706	1.13	.44	-43
Pembrolizumab	Platinum	NSCLC	KEYNOTE-042 ^{51d}	2019	1274	0.81	.002	12

(continued)

Table 1. SIFI of Overall Survival Calculated for 45 Phase 3 Trials Evaluating Immune Checkpoint Inhibitors in the Intention-to-Treat Populations (continued)

Intervention	Control	Tumor type	Clinical trial	Year	Sample size	HR	P value ^a	SIFI ^b
Anti-PD-L1								
Atezolizumab	Paclitaxel, docetaxel, or vinflunine	UC	IMvigor211 ⁵²	2018	931	0.85	.02	3
Atezolizumab	Docetaxel	NSCLC	OAK ^{53f}	2018	850	0.85	.0003	12
Atezolizumab + bevacizumab + carboplatin + paclitaxel	Bevacizumab + carboplatin + paclitaxel	Nonsquamous NSCLC	IMPpower150 ^{54g}	2018	696	0.78	.02	4
Atezolizumab + carboplatin + etoposide	Carboplatin + etoposide	SCLC	IMPpower133 ⁵⁵	2018	403	0.70	.01	3
Atezolizumab + carboplatin + nab-paclitaxel	Carboplatin + nab-paclitaxel	Nonsquamous NSCLC	IMPpower130 ⁵⁶	2019	723	0.79	.03	2
Atezolizumab + bevacizumab	Sunitinib	RCC	IMmotion151 ⁵⁷	2019	915	0.93	.71	-23
Atezolizumab + nab-paclitaxel	Nab-paclitaxel	BRCA	IMPpassion130 ⁷	2020	846	0.86	.13	-4
Atezolizumab	Regorafenib	CRC	IMblaze370 ⁵	2019	180	1.19	.35	-12
Atezolizumab + cobimetinib	Regorafenib	CRC	IMblaze370 ⁵	2019	273	1.0	.8	-9
Avelumab	Paclitaxel or irinotecan	GC or GEJC	JAVELIN Gastric 300 ⁵⁸	2018	371	1.1	.47	-13
Avelumab	Docetaxel	NSCLC	JAVELIN Lung 200 ^{59d}	2018	529	0.9	.36	-6
Durvalumab	Placebo	NSCLC	PACIFIC ⁶⁰	2018	713	0.68	.001	15
Durvalumab + platinum + etoposide	Platinum + etoposide	SCLC	CASPIAN ⁶¹	2019	537	0.73	.003	6
Anti-PD-1 + anti-CTLA-4								
Ipilimumab + nivolumab	Ipilimumab	Melanoma	CheckMate 067 ³⁶	2018	629	0.54	<.0001	38
Ipilimumab + nivolumab	Sunitinib	RCC	CheckMate 214 ⁴	2019	1096	0.71	.003	18
Ipilimumab + nivolumab	Platinum doublet	NSCLC	CheckMate 227 ⁶²	2019	1166	0.73	<.0001	24

Abbreviations: BRCA, breast cancer; CRC, colorectal cancer; CTLA-4, cytotoxic T-lymphocyte-associated protein 4; ESCC, esophageal squamous cell carcinoma; GC, gastric cancer; GEJC, gastroesophageal junction cancer; HCC, hepatocellular carcinoma; HNSCC, head and neck squamous cell carcinoma; HR, hazard ratio; MM, multiple myeloma; NSCLC, non-small cell lung carcinoma; PC, prostate cancer; PD-1, programmed cell death 1; PD-L1, programmed cell death 1 ligand 1; RCC, renal cell carcinoma; SCLC, small cell lung carcinoma; SIFI, survival-inferred fragility index; UC, urothelial carcinoma.

^a Calculated using 2-sided unstratified log-rank test.

^b Survival-inferred fragility index associated with the calculated P value (α = .05).

^c Modified intention-to-treat populations.

^d PD-L1 ≥ 1%.

^e PD-L1 ≥ 50%.

^f Intention-to-treat populations (n = 850).

^g EGFR or ALK wild-type.

populations and 16 (44%) of the 36 subgroup populations, the SIFI was less than the number of patients censored in the interventional group during only the first ventile (1/20th) of the follow-up time (eFigure 1 in the [Supplement](#)).

A comparison between positive SIFI levels in different tumor types among ITT populations (**Figure 2**) showed that non-small cell lung carcinoma, renal cell carcinoma, and melanoma had the highest values and that hepatocellular carcinoma, head and neck squamous cell carcinoma, and small cell lung carcinoma had the lowest values. Examining the association between SIFI and *P* values (in logarithmic scale) revealed a high correlation in ITT populations ($R = 0.70$; $P < 1 \times 10^{-7}$) and subgroup populations ($R = 0.82$; $P < 1 \times 10^{-9}$). However, the level of SIFI was not explained entirely by the variation in *P* values. For example, despite having relatively similar *P* values, hazard ratios, and sample sizes, the SIFI was 2-fold higher in KEYNOTE-024⁴⁰ compared with IMpower133,⁵⁵ and in ATTRACTION-2³³ compared with CheckMate 067³⁶ monotherapy (Table 1,^{2,7,24-62} **Figure 3**), indicating higher robustness. These examples demonstrate that statistical significance depends on the distribution of the longest-surviving patients, with more fragile studies relying on fewer patients to drive the significance, compared with less fragile studies that are associated with a higher "reserve" of patients. Similar associations between SIFI as a proportion of the population and *P* values are shown in eFigure 2 in the [Supplement](#). To explore the potential association of longer follow-up periods with the SIFI, we identified trials that published overall survival results for earlier follow-up periods. We found that the SIFI is stable and displays only a small variation for trials at different follow-up periods (**Table 2**),^{3,4,24,36,37,45,66-70} including studies with median follow-up time more than twice as long as in the original publication. Furthermore, we explored the operating characteristics of the SIFI, including sample size, censoring rate, and effect size (eFigures 3-5 in the [Supplement](#)). Performing simulations using combinations of the parameters resulted in 15 000 synthetic time-to-event data sets. Hazard ratios ranged from 0.13 to 1.95, and the percentage of individuals censored ranged from 17.5% to 50%. The simulated results provide a reference for the ranges of the SIFI for the various parameters of survival data.

The fragility for survival data can be calculated in various ways. Overall, we calculated 4 versions of SIFI, which include reassigning patients (flip) or adding patients (clone) to the opposite group using the best survivors from the experimental group or worst survivors from the control group. A comparison of the different SIFI approaches is shown for the ITT populations in eFigure 6 in the [Supplement](#). Compared with the default SIFI (flipping the best survivors to the opposite group) with a magnitude of 9 (IQR, 5-18) for ITT populations, the 3 alternative versions are associated with higher values in most studies. The SIFI magnitudes are 11 (IQR, 8-18) for flipping the worst survivors to the opposite group, 17.5 (IQR, 7-38.3) for cloning the best survivors to the opposite group, and 24 (IQR, 16-35) for cloning the worst survivors to the opposite group. These findings suggest that the SIFI using the version that flips the best survivors to the opposite group is the most sensitive approach for detecting the minimum changes required to overturn the conclusions.

Discussion

In our study, we found that the statistical significance of a substantial amount of phase 3 trials of ICIs could be lost or gained with a change in assignment of very few of the best surviving patients, often less than 1% of the respective trial sample size. Although this is an arbitrary number and does not reflect a random sampling of the patients, it represents a small fraction of the population that can overturn the statistical conclusions. Also, the change in the number of patients required for fragility is often smaller than the number of patients censored in the experimental group shortly after randomization, adding further uncertainties and raising concerns about the statistical outcomes had these and other patients been assessed to their end point. Eligibility for treatment with ICIs is assessed by concluding whether results of a trial are positive or negative. Our findings demonstrate how unstable these conclusions may be, and explain, in part, the widening gap between eligibility and benefit associated with ICIs.

The original fragility index has been applied to RCTs in oncology and other areas of medicine.^{17,19,71-74} However, the original fragility index is based on binary outcomes and the Fisher exact test, which could be misleading for time-to-event data, in which the primary interest is the

Figure 2. Survival-Inferred Fragility Index (SIFI) of Overall Survival in Phase 3 Randomized Clinical Trials

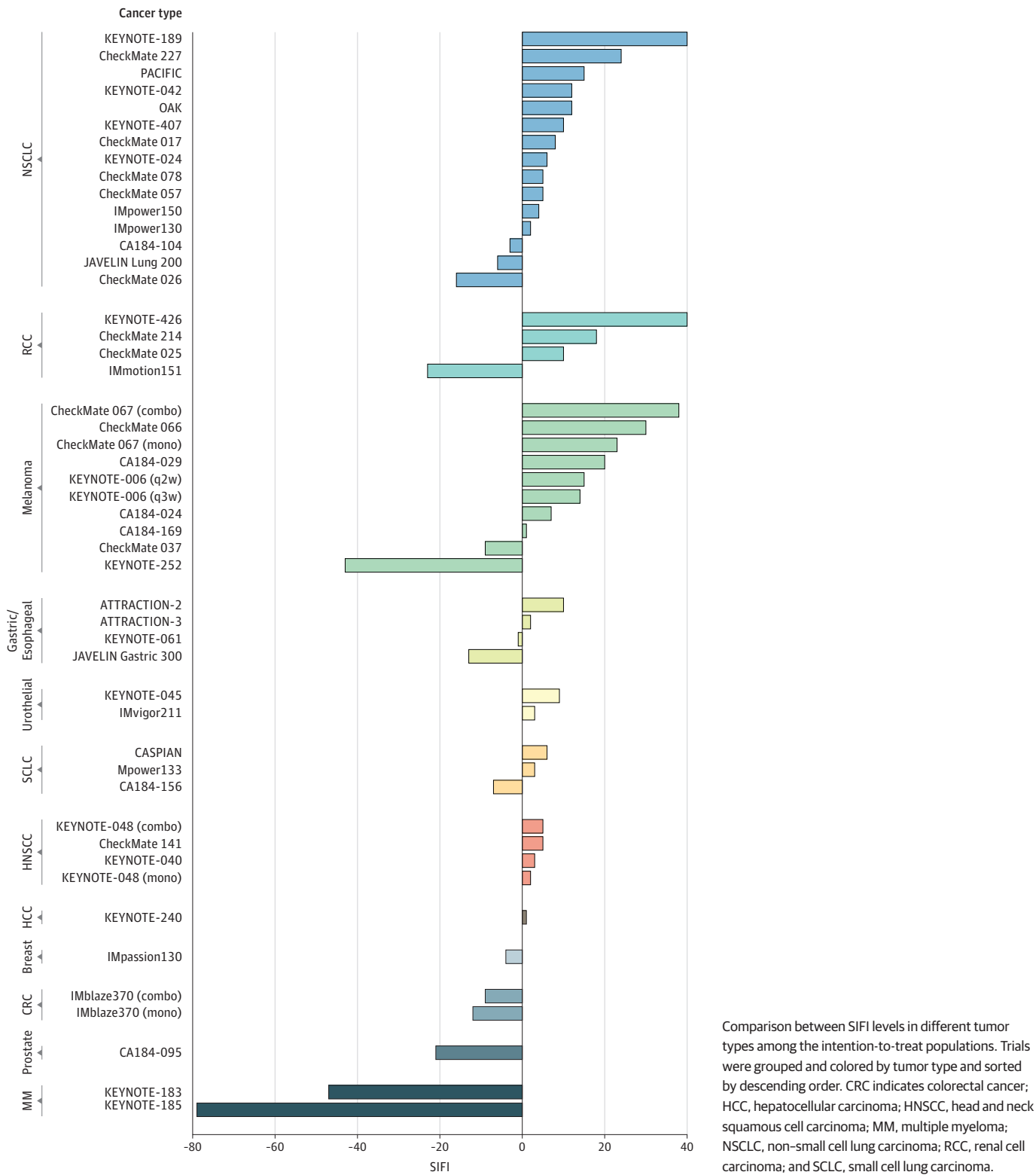
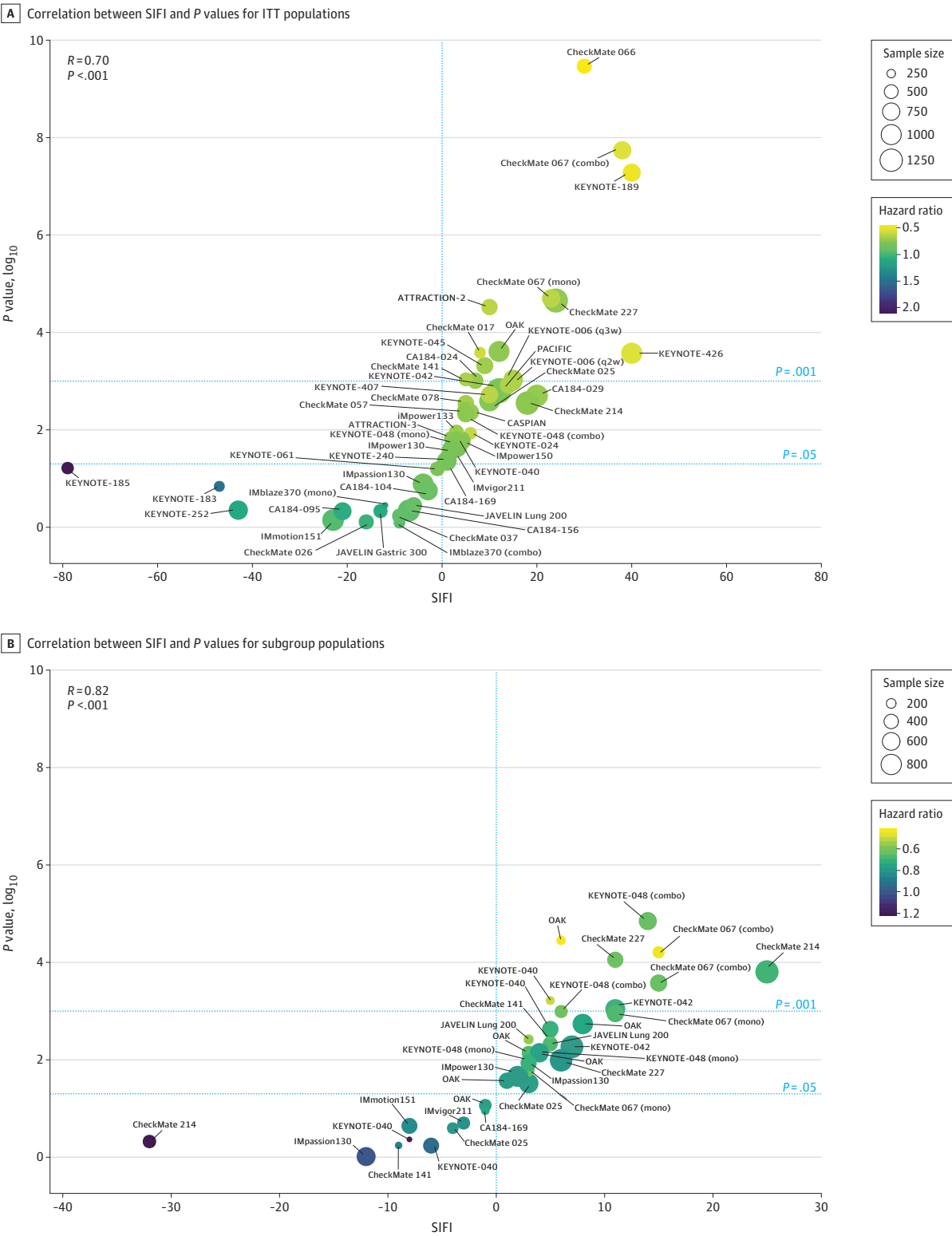


Figure 3. Survival-Inferred Fragility Index (SIFI) of Overall Survival in Phase 3 Randomized Clinical Trials



A, Correlation between SIFI and P values in a logarithmic scale for the intention-to-treat (ITT) populations. B, Correlation between SIFI and P values in a logarithmic scale for the subgroup populations. Color bars indicate hazard ratios and circle size represents the sample size. Correlation was calculated using Pearson correlation coefficient. Horizontal lines denoting .05 and .001 P value thresholds are shown.

timing of events.¹⁹ Although descriptions of time-to-event fragility exist,^{18,19} to our knowledge, no previous peer-reviewed original investigations have estimated time-aware fragility index for clinical trials, including oncology trials. Also, to our knowledge, no study has evaluated negative fragility measures for survival analysis.

In general, the *P* value serves as a measure of the compatibility of collected data with a defined statistical model. In a testing framework, smaller *P* values indicate greater evidence against the null hypothesis—a conjecture of no difference between outcomes of the intervention and control groups.⁷⁵ Undoubtedly, the *P* value plays a central role in the clinical testing of new drugs, and since the 1960s, the FDA has relied on significance testing to establish their effectiveness in the approval process.⁷⁶ As such, nowhere is this role more important than in clinical trials, where the smallest change in the *P* value can decisively influence the drug approval process and result in trial success or failure. Consequently, passing the statistical significance threshold has become the ultimate goal, and unless an analysis is adequately prespecified, most research designs allow enough leeway to manipulate the results to claim importance.⁷⁷⁻⁸⁰ Therefore, reliance on *P* values falling to either side of the significance threshold can result in extreme conclusions and be misleading, especially for a low threshold such as *P* < .05. Recently, an influential commentary published in *Nature*¹² has even called for the abandonment of the conventional threshold for statistical significance, regardless of the level (eg, *P* < .05), owing to this imposed dichotomization. However, statistical inferences are unavoidably dichotomous in many scientific fields. Most decisions in medicine are dichotomous, such as a new drug will either be approved or not, and will either be prescribed or not.⁷⁷

This study introduces the SIFI as a novel measure that enables us to estimate the vulnerability of the statistical conclusions of clinical trials with time-to-event outcomes. This index transforms the dichotomous conclusion to a discrete variable that provides more perspective regarding the potential benefit associated with ICIs or any other intervention. The SIFI provides context to the *P* value and statistical significance, which may not necessarily be intuitive and are often poorly understood.⁷⁷ Therefore, the SIFI translates uncertainty to a specified number that represents actual patients and events and places it on a linear scale that allows for assessment of the robustness of the results. For example, consider 2 comparable studies with similar *P* values. Although the SIFI is not a measure of effect, a trial with a high SIFI with an acceptable association with the sample size and censoring provides more robustness than a trial with a small SIFI representing a small fraction of the

Table 2. Comparison of SIFI of Overall Survival Calculated for Trials in Different Follow-up Periods

Study	Tumor	Sample size	Publication year	Follow-up, median, mo	No. of events	SIFI ^a
CA184-024	Melanoma	502	2011 ⁶⁵	39.0	409	5
			2015 ²⁴	63.1	427	7
CheckMate-025	RCC	821	2015 ⁶⁶	22.2	397	8
			2019 ⁴	45.6	567	10
CheckMate-067						
Combination therapy	Melanoma	629	2017 ⁶⁷	40.2	346	37
			2018 ³⁶	51.5	364	38
Monotherapy	Melanoma	631	2017 ⁶⁷	40.2	364	23
			2018 ³⁶	51.5	384	23
CheckMate-141	HNSCC	361	2018 ⁶⁸	17.8	288	4
			2018 ³⁷	30.6	321	5
KEYNOTE-006						
Every 2 wk	Melanoma	557	2015 ⁶⁹	13.3	196	18
			2017 ³	22.5	262	15
Every 3 wk	Melanoma	555	2015 ⁶⁹	13.3	203	12
			2017 ³	22.5	260	14
KEYNOTE-045	UC	542	2017 ⁷⁰	13.6	333	8
			2019 ⁴⁵	28.2	423	9

Abbreviations: HNSCC, head and neck squamous cell carcinoma; RCC, renal cell carcinoma; SIFI, survival-inferred fragility index; UC, urothelial carcinoma.

^a SIFI associated with the calculated *P* value (α = .05).

sample size and censoring. The latter relies on fragile evidence with higher uncertainty regarding the incompatibility with the null hypothesis. We did not define criteria for fragile vs nonfragile values, nor do we believe that a measure aimed to address the dichotomization of results by a threshold should be replaced by another. Perhaps trials involving the addition of a costly and a toxic drug to the standard treatment with a small effect size would require a higher level of robustness than trials comparing 2 drugs with similar overall properties. In contrast, concluding that statistically significant results show no real association when the fragility measure is very low is discouraged; it is equally inaccurate to claim that nonsignificant results with very small negative fragility point to an important signal. However, the SIFI allows for putting these 2 scenarios in context, expressing uncertainty and suggesting that the interpretation of their importance should be similar or, de facto, the same. In both cases, and especially for negative fragility measures, small values indicate that the true underlying effects either are negligible or lack statistical power. Nevertheless, considerations such as study design, data quality, comprehension of the underlying mechanisms, and other factors may often have more importance than statistical findings¹² such as *P* values or fragility indices.

The default solution for improving the confidence level would be making the barrier more demanding; however, this is a suboptimal option because the chance for false-negative results increases accordingly, and it still fails to address the vulnerability of the statistics. Nevertheless, fragility corresponding to one threshold is not comparable with another, and it is reasonable to expect lower fragility measures for lower *P* value thresholds, as they are interrelated. Hence, the approach encourages using lower significance thresholds. A trial not meeting a low prespecified significance threshold (eg, $P < .0001$), with a small negative SIFI (eg, -2), may provide higher confidence in the validity of the results compared with a trial that meets a higher threshold (eg, $P < .05$) but has a low positive SIFI (eg, 2). The SIFI relative to sample size can be useful to estimate the robustness of the results, but it could be misleading for small sample sizes. Although SIFI less than 1% in many RCTs could suggest extreme fragility, small trials with less than 100 patients cannot achieve a SIFI of less than 1%, even when the results are certainly less robust. Therefore, the SIFI relative to sample size, especially for small trials, should not be interpreted alone and must be accompanied by the SIFI.

Limitations

Several limitations of the study should be recognized. We did not address prespecified *P* value thresholds, which were allocated and controlled differently in every trial and are often much lower than .05. Instead, we used the standard α level of .05 as a common reference; therefore, some trials did not meet the prespecified threshold but resulted in a positive SIFI. Although not a strict rule by the FDA, the standard 2-trial α level is .05 but is smaller for approval based on a single trial.⁷⁶ The analysis of overall survival was based on an unstratified log-rank test at a 2-sided significance level as a uniform statistical test for all trials; however, studies have analyzed the data differently (eg, stratified or weighted log-rank test). Therefore, small differences exist between the published *P* value and the calculated *P* value. Furthermore, we found a small discrepancy in the numbers of patients at risk published in the original publications and the reproduced curves. For 19 of the 49 populations in the trials (39%), there was no discrepancy between the published and estimated number at risk at any time point. In the time points for which discrepancy existed, we found the difference to be small, with a median of 1 patient (IQR, 1-2).

The SIFI can be calculated in various ways. Our comparison of different implementations of the SIFI demonstrates that reassigning or adding the best survivors to the opposite group provides lower fragility estimates compared with the worst survivors, for most trials. This finding indicates that the longest-surviving patients can tilt the balance between the groups more strongly compared with the shortest-surviving patients. The association of the longest survivors with the survival curves is potentially unlimited, as they are constrained only by the follow-up time, whereas the shortest-surviving patients cannot have an event before time zero. By both removing a long-time survivor from one group and adding them to the other group, the total number of patients required to pass

the significance threshold is reduced compared with other techniques. This approach coincides with the essence of fragility—identifying the minimum required changes to overturn the conclusions. Furthermore, we aimed to define a simple and intuitive method that can be recreated using existing routines, is quantifiable in all conditions, and is applicable to real-world practice in which patients are randomly assigned from a pool of eligible patients. Although random variations alone can lead to large disparities in *P* values, the calculation of the SIFI is not based on random variations in the assignment of patients but on the reassignment of patients at the extreme ends of the scale. However, the random allocation of patients can lead to different proportions of the best (or worst) survivors in the groups, which may impact the outcomes. Therefore, the SIFI serves as a simple and conservative approach to reflect the fragility of the statistics. Alternatively, the mean or median survival time can be exploited in different ways to quantify the fragility^{18,19}; however, this approach can underestimate the fragility if the few patients who cause most of the difference are not captured.

Conclusions

The results of this study suggest that many phase 3 RCTs evaluating ICI therapies are fragile and challenge the confidence in rejecting or concluding superiority for these drugs compared with standard treatments. Low fragility levels express uncertainty when there is no appreciable difference between the interpretative significance of data. In contrast, high fragility levels can provide robustness and aid in binary decision-making, especially for treatments associated with high cost and toxic effects that require strong support. Interpretation of any outcome is far more complicated than just significance testing, and the SIFI as a statistical and communication tool may serve as a better starting point for discerning between science and fiction.

ARTICLE INFORMATION

Accepted for Publication: July 13, 2020.

Published: October 23, 2020. doi:10.1001/jamanetworkopen.2020.17675

Open Access: This is an open access article distributed under the terms of the [CC-BY License](#). © 2020 Bomze D et al. JAMA Network Open.

Corresponding Authors: Gal Markel, MD, PhD (gal.markel@sheba.health.gov.il), and Tomer Meirson, BSc (tomermrsn@gmail.com), Ella Lemelbaum Institute for Immuno-Oncology, Sheba Medical Center, Ramat-Gan 526260, Israel.

Author Affiliations: Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel (Bomze); Institute for Immunobiology, Kantonsspital St Gallen, St Gallen, Switzerland (Bomze, Hasan Ali, Flatz); Ella Lemelbaum Institute for Immuno-Oncology, Sheba Medical Center, Ramat-Gan, Israel (Asher, Markel, Meirson); Department of Dermatology, University Hospital of Zurich, Zurich, Switzerland (Hasan Ali, Flatz); Department of Oncology, Kantonsspital St Gallen, St Gallen, Switzerland (Flatz); Center for Liver Diseases, Sheba Medical Center, Ramat-Gan, Israel (Azoulay); Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel (Markel); Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel (Meirson).

Author Contributions: Messrs Bomze and Meirson had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Bomze, Hasan Ali, Azoulay, Markel, Meirson.

Acquisition, analysis, or interpretation of data: Bomze, Asher, Flatz, Meirson.

Drafting of the manuscript: Bomze, Meirson.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Bomze, Meirson.

Administrative, technical, or material support: Asher, Meirson.

Supervision: Bomze, Azoulay, Markel, Meirson.

Conflict of Interest Disclosures: Dr Asher reported receiving personal fees from MSD, BMS, Medison, and Novartis outside the submitted work. Dr Flatz reported receiving grants from Swiss National Science Foundation,

Swiss Cancer League, Hookipa Pharma, and Novartis Foundation outside the submitted work. Dr Markel reported receiving personal fees from MSD and Roche; grants and personal fees from BMS and Novartis; personal fees and stock options from 4C Biomed; and stock options from Nucleai, Biond Biologics, and Ella Therapeutics outside the submitted work. Mr Meirson reported receiving a grant from the Foulkes Foundation for MD/PhD students. No other disclosures were reported.

Funding/Support: Dr Platz is supported by a Swiss National Science Foundation professorship (PPOOP3_157448). Dr Markel is supported by the Samulei Foundation Grant for Integrative Immuno-Oncology. Mr Meirson is supported by the Foulkes Foundation fellowship for MD/PhD students.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Smyth MJ, Ngiew SF, Ribas A, Teng MW. Combination cancer immunotherapies tailored to the tumour microenvironment. *Nat Rev Clin Oncol*. 2016;13(3):143-158. doi:10.1038/nrclinonc.2015.209
2. Usmani SZ, Schjesvold F, Oriol A, et al; KEYNOTE-185 Investigators. Pembrolizumab plus lenalidomide and dexamethasone for patients with treatment-naïve multiple myeloma (KEYNOTE-185): a randomised, open-label, phase 3 trial. *Lancet Haematol*. 2019;6(9):e448-e458. doi:10.1016/S2352-3026(19)30109-7
3. Schachter J, Ribas A, Long GV, et al. Pembrolizumab versus ipilimumab for advanced melanoma: final overall survival results of a multicentre, randomised, open-label phase 3 study (KEYNOTE-006). *Lancet*. 2017;390(10105):1853-1862. doi:10.1016/S0140-6736(17)31601-X
4. Motzer RJ, Rini BI, McDermott DF, et al; CheckMate 214 investigators. Nivolumab plus ipilimumab versus sunitinib in first-line treatment for advanced renal cell carcinoma: extended follow-up of efficacy and safety results from a randomised, controlled, phase 3 trial. *Lancet Oncol*. 2019;20(10):1370-1385. doi:10.1016/S1470-2045(19)30413-9
5. Eng C, Kim TW, Bendell J, et al; IMblaze370 Investigators. Atezolizumab with or without cobimetinib versus regorafenib in previously treated metastatic colorectal cancer (IMblaze370): a multicentre, open-label, phase 3, randomised, controlled trial. *Lancet Oncol*. 2019;20(6):849-861. doi:10.1016/S1470-2045(19)30027-0
6. Beer TM, Kwon ED, Drake CG, et al. Randomized, double-blind, phase III trial of ipilimumab versus placebo in asymptomatic or minimally symptomatic patients with metastatic chemotherapy-naïve castration-resistant prostate cancer. *J Clin Oncol*. 2017;35(1):40-47. doi:10.1200/JCO.2016.69.1584
7. Schmid P, Rugo HS, Adams S, et al; IMpassion130 Investigators. Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol*. 2020;21(1):44-59. doi:10.1016/S1470-2045(19)30689-8
8. Beaver JA, Howie LJ, Pelosof L, et al. A 25-year experience of US Food and Drug Administration accelerated approval of malignant hematology and oncology drugs and biologics: a review. *JAMA Oncol*. 2018;4(6):849-856. doi:10.1001/jamaoncol.2017.5618
9. Gill J, Prasad V. A reality check of the accelerated approval of immune-checkpoint inhibitors. *Nat Rev Clin Oncol*. 2019;16(11):656-658. doi:10.1038/s41571-019-0260-y
10. Haslam A, Gill J, Prasad V. Estimation of the percentage of US patients with cancer who are eligible for immune checkpoint inhibitor drugs. *JAMA Netw Open*. 2020;3(3):e200423. doi:10.1001/jamanetworkopen.2020.0423
11. Catenacci DVT, Hochster H, Klempner SJ. Keeping checkpoint inhibitors in check. *JAMA Netw Open*. 2019;2(5):e192546. doi:10.1001/jamanetworkopen.2019.2546
12. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature*. 2019;567(7748):305-307. doi:10.1038/d41586-019-00857-9
13. Goodman SN. Toward evidence-based medical statistics, 1: the P value fallacy. *Ann Intern Med*. 1999;130(12):995-1004. doi:10.7326/0003-4819-130-12-199906150-00008
14. Del Paggio JC, Sullivan R, Schrag D, et al. Delivery of meaningful cancer care: a retrospective cohort study assessing cost and benefit with the ASCO and ESMO frameworks. *Lancet Oncol*. 2017;18(7):887-894. doi:10.1016/S1470-2045(17)30415-1
15. Cherny NI, Dafni U, Bogaerts J, et al. ESMO-magnitude of clinical benefit scale version 1.1. *Ann Oncol*. 2017;28(10):2340-2366. doi:10.1093/annonc/mdx310
16. Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a fragility index. *J Clin Epidemiol*. 2014;67(6):622-628. doi:10.1016/j.jclinepi.2013.10.019

17. Del Paggio JC, Tannock IF. The fragility of phase 3 trials supporting FDA-approved anticancer medicines: a retrospective analysis. *Lancet Oncol*. 2019;20(8):1065-1069. doi:[10.1016/S1470-2045\(19\)30338-9](https://doi.org/10.1016/S1470-2045(19)30338-9)
18. Johnson KW, Rappaport E, Shameer K, Glicksberg BS, Dudley JT. fragilityindex: an R package for statistical fragility estimates in biomedicine. Preprint. Posted online February 27, 2019. bioRxiv 562264. doi:[10.1101/562264](https://doi.org/10.1101/562264)
19. Bomze D, Meirson T. A critique of the fragility index. *Lancet Oncol*. 2019;20(10):e551. doi:[10.1016/S1470-2045\(19\)30582-0](https://doi.org/10.1016/S1470-2045(19)30582-0)
20. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg*. 2014;12(12):1495-1499. doi:[10.1016/j.ijsu.2014.07.013](https://doi.org/10.1016/j.ijsu.2014.07.013)
21. Wei Y, Royston P. Reconstructing time-to-event data from published Kaplan-Meier curves. *Stata J*. 2017;17(4):786-802. doi:[10.1177/1536867X1801700402](https://doi.org/10.1177/1536867X1801700402)
22. GitHub. Code for calculating the survival-inferred fragility index (SIFI). 2020. Accessed September 17, 2020. <https://github.com/davidbomze/SIFI>.
23. Moriña D, Navarro A. The R package survsim for the simulation of simple and complex survival data. *J Stat Software* 2014;59(2):1-20. doi:[10.18637/jss.v059.i02](https://doi.org/10.18637/jss.v059.i02)
24. Maio M, Grob JJ, Aamdal S, et al. Five-year survival rates for treatment-naïve patients with advanced melanoma who received ipilimumab plus dacarbazine in a phase III trial. *J Clin Oncol*. 2015;33(10):1191-1196. doi:[10.1200/JCO.2014.56.6018](https://doi.org/10.1200/JCO.2014.56.6018)
25. Eggermont AM, Chiarion-Sileni V, Grob JJ, et al. Prolonged survival in stage III melanoma with ipilimumab adjuvant therapy. *N Engl J Med*. 2016;375(19):1845-1855. doi:[10.1056/NEJMoa1611299](https://doi.org/10.1056/NEJMoa1611299)
26. Reck M, Luft A, Szczesna A, et al. Phase III randomized trial of ipilimumab plus etoposide and platinum versus placebo plus etoposide and platinum in extensive-stage small-cell lung cancer. *J Clin Oncol*. 2016;34(31):3740-3748. doi:[10.1200/JCO.2016.67.6601](https://doi.org/10.1200/JCO.2016.67.6601)
27. Govindan R, Szczesna A, Ahn MJ, et al. Phase III trial of ipilimumab combined with paclitaxel and carboplatin in advanced squamous non-small-cell lung cancer. *J Clin Oncol*. 2017;35(30):3449-3457. doi:[10.1200/JCO.2016.71.7629](https://doi.org/10.1200/JCO.2016.71.7629)
28. Ascierto PA, Del Vecchio M, Robert C, et al. Ipilimumab 10 mg/kg versus ipilimumab 3 mg/kg in patients with unresectable or metastatic melanoma: a randomised, double-blind, multicentre, phase 3 trial. *Lancet Oncol*. 2017;18(5):611-622. doi:[10.1016/S1470-2045\(17\)30231-0](https://doi.org/10.1016/S1470-2045(17)30231-0)
29. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med*. 2015;373(2):123-135. doi:[10.1056/NEJMoa1504627](https://doi.org/10.1056/NEJMoa1504627)
30. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med*. 2015;373(17):1627-1639. doi:[10.1056/NEJMoa1507643](https://doi.org/10.1056/NEJMoa1507643)
31. Tomita Y, Fukasawa S, Shinohara N, et al. Nivolumab versus everolimus in advanced renal cell carcinoma: Japanese subgroup 3-year follow-up analysis from the phase III CheckMate 025 study. *Jpn J Clin Oncol*. 2019;49(6):506-514. doi:[10.1093/jjco/hyz026](https://doi.org/10.1093/jjco/hyz026)
32. Carbone DP, Reck M, Paz-Ares L, et al; CheckMate 026 Investigators. First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. *N Engl J Med*. 2017;376(25):2415-2426. doi:[10.1056/NEJMoa1613493](https://doi.org/10.1056/NEJMoa1613493)
33. Kang YK, Boku N, Satoh T, et al. Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2017;390(10111):2461-2471. doi:[10.1016/S0140-6736\(17\)31827-5](https://doi.org/10.1016/S0140-6736(17)31827-5)
34. Larkin J, Minor D, D'Angelo S, et al. Overall survival in patients with advanced melanoma who received nivolumab versus investigator's choice chemotherapy in CheckMate 037: a randomized, controlled, open-label phase III trial. *J Clin Oncol*. 2018;36(4):383-390. doi:[10.1200/JCO.2016.71.8023](https://doi.org/10.1200/JCO.2016.71.8023)
35. Ascierto PA, Long GV, Robert C, et al. Survival outcomes in patients with previously untreated *BRAF* wild-type advanced melanoma treated with nivolumab therapy: three-year follow-up of a randomized phase 3 trial. *JAMA Oncol*. 2019;5(2):187-194. doi:[10.1001/jamaoncol.2018.4514](https://doi.org/10.1001/jamaoncol.2018.4514)
36. Hodi FS, Chiarion-Sileni V, Gonzalez R, et al. Nivolumab plus ipilimumab or nivolumab alone versus ipilimumab alone in advanced melanoma (CheckMate 067): 4-year outcomes of a multicentre, randomised, phase 3 trial. *Lancet Oncol*. 2018;19(11):1480-1492. doi:[10.1016/S1470-2045\(18\)30700-9](https://doi.org/10.1016/S1470-2045(18)30700-9)
37. Ferris RL, Blumenschein G Jr, Fayette J, et al. Nivolumab vs investigator's choice in recurrent or metastatic squamous cell carcinoma of the head and neck: 2-year long-term survival update of CheckMate 141 with analyses by tumor PD-L1 expression. *Oral Oncol*. 2018;81:45-51. doi:[10.1016/j.oraloncology.2018.04.008](https://doi.org/10.1016/j.oraloncology.2018.04.008)

38. Kato K, Cho BC, Takahashi M, et al. Nivolumab versus chemotherapy in patients with advanced oesophageal squamous cell carcinoma refractory or intolerant to previous chemotherapy (ATTRACTION-3): a multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol*. 2019;20(11):1506-1517. doi:[10.1016/S1470-2045\(19\)30626-6](https://doi.org/10.1016/S1470-2045(19)30626-6)
39. Wu YL, Lu S, Cheng Y, et al. Nivolumab versus docetaxel in a predominantly Chinese patient population with previously treated advanced NSCLC: CheckMate 078 randomized phase III clinical trial. *J Thorac Oncol*. 2019;14(5):867-875. doi:[10.1016/j.jtho.2019.01.006](https://doi.org/10.1016/j.jtho.2019.01.006)
40. Reck M, Rodríguez-Abreu D, Robinson AG, et al; KEYNOTE-024 Investigators. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med*. 2016;375(19):1823-1833. doi:[10.1056/NEJMoa1606774](https://doi.org/10.1056/NEJMoa1606774)
41. Cohen EEW, Soulières D, Le Tourneau C, et al; KEYNOTE-040 investigators. Pembrolizumab versus methotrexate, docetaxel, or cetuximab for recurrent or metastatic head-and-neck squamous cell carcinoma (KEYNOTE-040): a randomised, open-label, phase 3 study. *Lancet*. 2019;393(10167):156-167. doi:[10.1016/S0140-6736\(18\)31999-8](https://doi.org/10.1016/S0140-6736(18)31999-8)
42. Shitara K, Özgüroğlu M, Bang YJ, et al; KEYNOTE-061 investigators. Pembrolizumab versus paclitaxel for previously treated, advanced gastric or gastro-oesophageal junction cancer (KEYNOTE-061): a randomised, open-label, controlled, phase 3 trial. *Lancet*. 2018;392(10142):123-133. doi:[10.1016/S0140-6736\(18\)31257-1](https://doi.org/10.1016/S0140-6736(18)31257-1)
43. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al; KEYNOTE-189 Investigators. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N Engl J Med*. 2018;378(22):2078-2092. doi:[10.1056/NEJMoa1801005](https://doi.org/10.1056/NEJMoa1801005)
44. Paz-Ares L, Luft A, Vicente D, et al; KEYNOTE-407 Investigators. Pembrolizumab plus chemotherapy for squamous non-small-cell lung cancer. *N Engl J Med*. 2018;379(21):2040-2051. doi:[10.1056/NEJMoa1810865](https://doi.org/10.1056/NEJMoa1810865)
45. Fradet Y, Bellmunt J, Vaughn DJ, et al. Randomized phase III KEYNOTE-045 trial of pembrolizumab versus paclitaxel, docetaxel, or vinflunine in recurrent advanced urothelial cancer: results of >2 years of follow-up. *Ann Oncol*. 2019;30(6):970-976. doi:[10.1093/annonc/mdz127](https://doi.org/10.1093/annonc/mdz127)
46. Burtneß B, Harrington KJ, Greil R, et al; KEYNOTE-048 Investigators. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. *Lancet*. 2019;394(10212):1915-1928. doi:[10.1016/S0140-6736\(19\)32591-7](https://doi.org/10.1016/S0140-6736(19)32591-7)
47. Mateos MV, Blacklock H, Schjesvold F, et al; KEYNOTE-183 Investigators. Pembrolizumab plus pomalidomide and dexamethasone for patients with relapsed or refractory multiple myeloma (KEYNOTE-183): a randomised, open-label, phase 3 trial. *Lancet Haematol*. 2019;6(9):e459-e469. doi:[10.1016/S2352-3026\(19\)30110-3](https://doi.org/10.1016/S2352-3026(19)30110-3)
48. Finn RS, Ryoo BY, Merle P, et al; KEYNOTE-240 investigators. Pembrolizumab as second-line therapy in patients with advanced hepatocellular carcinoma in KEYNOTE-240: a randomized, double-blind, phase III trial. *J Clin Oncol*. 2020;38(3):193-202. doi:[10.1200/JCO.19.01307](https://doi.org/10.1200/JCO.19.01307)
49. Rini BI, Plimack ER, Stus V, et al; KEYNOTE-426 Investigators. Pembrolizumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. *N Engl J Med*. 2019;380(12):1116-1127. doi:[10.1056/NEJMoa1816714](https://doi.org/10.1056/NEJMoa1816714)
50. Long GV, Dummer R, Hamid O, et al. Epcadostat plus pembrolizumab versus placebo plus pembrolizumab in patients with unresectable or metastatic melanoma (ECHO-301/KEYNOTE-252): a phase 3, randomised, double-blind study. *Lancet Oncol*. 2019;20(8):1083-1097. doi:[10.1016/S1470-2045\(19\)30274-8](https://doi.org/10.1016/S1470-2045(19)30274-8)
51. Mok TSK, Wu YL, Kudaba I, et al; KEYNOTE-042 Investigators. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet*. 2019;393(10183):1819-1830. doi:[10.1016/S0140-6736\(18\)32409-7](https://doi.org/10.1016/S0140-6736(18)32409-7)
52. Powles T, Durán I, van der Heijden MS, et al. Atezolizumab versus chemotherapy in patients with platinum-treated locally advanced or metastatic urothelial carcinoma (IMvigor211): a multicentre, open-label, phase 3 randomised controlled trial. *Lancet*. 2018;391(10122):748-757. doi:[10.1016/S0140-6736\(17\)33297-X](https://doi.org/10.1016/S0140-6736(17)33297-X)
53. Fehrenbacher L, von Pawel J, Park K, et al. Updated efficacy analysis including secondary population results for OAK: a randomized phase III study of atezolizumab versus docetaxel in patients with previously treated advanced non-small cell lung cancer. *J Thorac Oncol*. 2018;13(8):1156-1170. doi:[10.1016/j.jtho.2018.04.039](https://doi.org/10.1016/j.jtho.2018.04.039)
54. Socinski MA, Jotte RM, Cappuzzo F, et al; IMPower150 Study Group. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. *N Engl J Med*. 2018;378(24):2288-2301. doi:[10.1056/NEJMoa1716948](https://doi.org/10.1056/NEJMoa1716948)
55. Horn L, Mansfield AS, Szczesna A, et al; IMPower133 Study Group. First-line atezolizumab plus chemotherapy in extensive-stage small-cell lung cancer. *N Engl J Med*. 2018;379(23):2220-2229. doi:[10.1056/NEJMoa1809064](https://doi.org/10.1056/NEJMoa1809064)

56. West H, McCleod M, Hussein M, et al. Atezolizumab in combination with carboplatin plus nab-paclitaxel chemotherapy compared with chemotherapy alone as first-line treatment for metastatic non-squamous non-small-cell lung cancer (IMpower130): a multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol*. 2019;20(7):924-937. doi:[10.1016/S1470-2045\(19\)30167-6](https://doi.org/10.1016/S1470-2045(19)30167-6)
57. Rini BI, Powles T, Atkins MB, et al; IMmotion151 Study Group. Atezolizumab plus bevacizumab versus sunitinib in patients with previously untreated metastatic renal cell carcinoma (IMmotion151): a multicentre, open-label, phase 3, randomised controlled trial. *Lancet*. 2019;393(10189):2404-2415. doi:[10.1016/S0140-6736\(19\)30723-8](https://doi.org/10.1016/S0140-6736(19)30723-8)
58. Bang YJ, Ruiz EY, Van Cutsem E, et al. Phase III, randomised trial of avelumab versus physician's choice of chemotherapy as third-line treatment of patients with advanced gastric or gastro-oesophageal junction cancer: primary analysis of JAVELIN Gastric 300. *Ann Oncol*. 2018;29(10):2052-2060. doi:[10.1093/annonc/mdy264](https://doi.org/10.1093/annonc/mdy264)
59. Barlesi F, Vansteenkiste J, Spigel D, et al. Avelumab versus docetaxel in patients with platinum-treated advanced non-small-cell lung cancer (JAVELIN Lung 200): an open-label, randomised, phase 3 study. *Lancet Oncol*. 2018;19(11):1468-1479. doi:[10.1016/S1470-2045\(18\)30673-9](https://doi.org/10.1016/S1470-2045(18)30673-9)
60. Antonia SJ, Villegas A, Daniel D, et al; PACIFIC Investigators. Overall survival with durvalumab after chemoradiotherapy in stage III NSCLC. *N Engl J Med*. 2018;379(24):2342-2350. doi:[10.1056/NEJMoa1809697](https://doi.org/10.1056/NEJMoa1809697)
61. Paz-Ares L, Dvorkin M, Chen Y, et al; CASPIAN investigators. Durvalumab plus platinum-etoposide versus platinum-etoposide in first-line treatment of extensive-stage small-cell lung cancer (CASPIAN): a randomised, controlled, open-label, phase 3 trial. *Lancet*. 2019;394(10212):1929-1939. doi:[10.1016/S0140-6736\(19\)32222-6](https://doi.org/10.1016/S0140-6736(19)32222-6)
62. Hellmann MD, Paz-Ares L, Bernabe Caro R, et al. Nivolumab plus ipilimumab in advanced non-small-cell lung cancer. *N Engl J Med*. 2019;381(21):2020-2031. doi:[10.1056/NEJMoa1910231](https://doi.org/10.1056/NEJMoa1910231)
63. Hodi FS, O'Day SJ, McDermott DF, et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med*. 2010;363(8):711-723. doi:[10.1056/NEJMoa1003466](https://doi.org/10.1056/NEJMoa1003466)
64. Kwon ED, Drake CG, Scher HI, et al; CA184-043 Investigators. Ipilimumab versus placebo after radiotherapy in patients with metastatic castration-resistant prostate cancer that had progressed after docetaxel chemotherapy (CA184-043): a multicentre, randomised, double-blind, phase 3 trial. *Lancet Oncol*. 2014;15(7):700-712. doi:[10.1016/S1470-2045\(14\)70189-5](https://doi.org/10.1016/S1470-2045(14)70189-5)
65. Robert C, Thomas L, Bondarenko I, et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N Engl J Med*. 2011;364(26):2517-2526. doi:[10.1056/NEJMoa1104621](https://doi.org/10.1056/NEJMoa1104621)
66. Motzer RJ, Escudier B, McDermott DF, et al; CheckMate 025 investigators. Nivolumab versus everolimus in advanced renal-cell carcinoma. *N Engl J Med*. 2015;373(19):1803-1813. doi:[10.1056/NEJMoa1510665](https://doi.org/10.1056/NEJMoa1510665)
67. Wolchok JD, Chiarion-Sileni V, Gonzalez R, et al. Overall survival with combined nivolumab and ipilimumab in advanced melanoma. *N Engl J Med*. 2017;377(14):1345-1356. doi:[10.1056/NEJMoa1709684](https://doi.org/10.1056/NEJMoa1709684)
68. Gillison ML, Blumenschein G Jr, Fayette J, et al. CheckMate 141: 1-year update and subgroup analysis of nivolumab as first-line therapy in patients with recurrent/metastatic head and neck cancer. *Oncologist*. 2018;23(9):1079-1082. doi:[10.1634/theoncologist.2017-0674](https://doi.org/10.1634/theoncologist.2017-0674)
69. Robert C, Schachter J, Long GV, et al; KEYNOTE-006 investigators. Pembrolizumab versus ipilimumab in advanced melanoma. *N Engl J Med*. 2015;372(26):2521-2532. doi:[10.1056/NEJMoa1503093](https://doi.org/10.1056/NEJMoa1503093)
70. Bellmunt J, de Wit R, Vaughn DJ, et al; KEYNOTE-045 Investigators. Pembrolizumab as second-line therapy for advanced urothelial carcinoma. *N Engl J Med*. 2017;376(11):1015-1026. doi:[10.1056/NEJMoa1613683](https://doi.org/10.1056/NEJMoa1613683)
71. Khan MS, Ochani RK, Shaikh A, et al. Fragility index in cardiovascular randomized controlled trials. *Circ Cardiovasc Qual Outcomes*. 2019;12(12):e005755. doi:[10.1161/CIRCOUTCOMES.119.005755](https://doi.org/10.1161/CIRCOUTCOMES.119.005755)
72. Gaudino M, Hameed I, Biondi-Zoccai G, et al. Systematic evaluation of the robustness of the evidence supporting current guidelines on myocardial revascularization using the fragility index. *Circ Cardiovasc Qual Outcomes*. 2019;12(12):e006017. doi:[10.1161/CIRCOUTCOMES.119.006017](https://doi.org/10.1161/CIRCOUTCOMES.119.006017)
73. Tignanelli CJ, Napolitano LM. The fragility index in randomized clinical trials as a means of optimizing patient care. *JAMA Surg*. 2019;154(1):74-79. doi:[10.1001/jamasurg.2018.4318](https://doi.org/10.1001/jamasurg.2018.4318)
74. Das S, Xavier S. Calculation of the fragility index of randomized controlled trials in epilepsy published in twelve major journals. *Epilepsy Res*. 2020;159:106258. doi:[10.1016/j.eplepsyres.2019.106258](https://doi.org/10.1016/j.eplepsyres.2019.106258)
75. Altman N, Krzywinski M. Points of significance: interpreting *P* values. *Nat Methods*. 2017;14(3):213-214. doi:[10.1038/nmeth.4210](https://doi.org/10.1038/nmeth.4210)
76. Kennedy-Shaffer L. When the alpha is the omega: *P*-values, "substantial evidence," and the 0.05 standard at FDA. *Food Drug Law J*. 2017;72(4):595-635.
77. Ioannidis JPA. The importance of predefined rules and prespecified statistical analyses: do not abandon significance. *JAMA*. 2019;321(21):2067-2068. doi:[10.1001/jama.2019.4582](https://doi.org/10.1001/jama.2019.4582)

78. Patel CJ, Burford B, Ioannidis JP. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *J Clin Epidemiol*. 2015;68(9):1046-1058. doi:10.1016/j.jclinepi.2015.05.029
79. Greenberg L, Jairath V, Pearce R, Kahan BC. Pre-specification of statistical analysis approaches in published clinical trial protocols was inadequate. *J Clin Epidemiol*. 2018;101:53-60. doi:10.1016/j.jclinepi.2018.05.023
80. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*. 2004;291(20):2457-2465. doi:10.1001/jama.291.20.2457

SUPPLEMENT.

eFigure 1. Association Between Survival-Inferred Fragility Index and Censoring of Phase 3 Randomized Clinical Trials

eFigure 2. Proportion of Survival-Inferred Fragility Index of Overall Survival in Phase 3 Randomized Clinical Trials

eFigure 3. Survival-Inferred Fragility Index for Synthetic Survival Data

eFigure 4. Relationship Between the Sample Size and Survival-Inferred Fragility Index for Synthetic Survival Data

eFigure 5. Relationship Between the Effects Size and Survival-Inferred Fragility Index for Synthetic Survival Data

eFigure 6. Comparison of Four Versions of the Survival-Inferred Fragility Index

eTable. Survival-Inferred Fragility Index of Overall Survival Calculated for 15 Phase 3 Trials Evaluating Immune Checkpoint Inhibitors in Subgroup Populations

eReferences.